

## Durham Research Online

---

### Deposited in DRO:

27 June 2014

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Rollinde, E. and Theuns, Tom and Schaye, J. and Pâris, I. and Petitjean, P. (2013) 'Sample variance and Lyman forest transmission statistics.', *Monthly notices of the Royal Astronomical Society.*, 428 (1). pp. 540-550.

### Further information on publisher's website:

<http://dx.doi.org/10.1093/mnras/sts057>

### Publisher's copyright statement:

This article has been accepted for publication in *Monthly notices of the Royal Astronomical Society* © 2013 The Authors Published by Oxford University Press on behalf of Royal Astronomical Society. All rights reserved.

### Additional information:

---

## Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Sample variance and Lyman $\alpha$ forest transmission statistics

E. Rollinde,<sup>1</sup>★ T. Theuns,<sup>2,3</sup> J. Schaye,<sup>4</sup> I. Pâris<sup>1</sup> and P. Petitjean<sup>1</sup>

<sup>1</sup>UPMC Université Paris 06, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France

<sup>2</sup>Institute of Computational Cosmology, Department of Physics, University of Durham, Science Laboratories, South Road, Durham DH1 3LE

<sup>3</sup>Universiteit Antwerpen, Campus Groenenborger, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

<sup>4</sup>Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands

Accepted 2012 September 20. Received 2012 September 20; in original form 2011 March 15

## ABSTRACT

We compare the observed probability distribution function (PDF) of the transmission in the H I Lyman  $\alpha$  forest, measured from the Ultraviolet and Visual Echelle Spectrograph (UVES) ‘Large Programme’ sample at redshifts  $z = [2, 2.5, 3]$ , to results from the GIMIC cosmological simulations. Our measured values for the mean transmission and its PDF are in good agreement with published results. Errors on statistics measured from high-resolution data are typically estimated using bootstrap or jackknife resampling techniques after splitting the spectra into chunks. We demonstrate that these methods tend to underestimate the sample variance unless the chunk size is much larger than is commonly the case. We therefore estimate the sample variance from the simulations. We conclude that observed and simulated transmission statistics are in good agreement; in particular, we do not require the temperature–density relation to be ‘inverted’.

**Key words:** methods: numerical – galaxies: intergalactic medium – cosmology: theory.

## 1 INTRODUCTION

At high redshift, the intergalactic medium (IGM) contains the majority of baryons in the Universe (Petitjean et al. 1993; Fukugita, Hogan & Peebles 1998) and is highly ionized by the ultraviolet background (UVB) produced by galaxies and quasi-stellar objects (QSOs; Gunn & Peterson 1965) at least since redshift  $z \sim 6$  (Fan, Carilli & Keating 2006; Becker, Rauch & Sargent 2007), becoming increasingly neutral near  $z \sim 7$  (Mortlock et al. 2011). It is detected in absorption against bright sources as the H I Lyman  $\alpha$  forest (Lynds 1971; see Rauch 1998, for a review).

High signal-to-noise ratio (S/N) observations with high-resolution, echelle spectrographs, such as the Ultraviolet and Visual Echelle Spectrograph (UVES) on the Very Large Telescope (VLT; e.g. Bergeron et al. 2004) and HIRES on Keck (e.g. Hu et al. 1995), of this forest of H I absorption lines together with numerical simulations (Cen et al. 1994; Petitjean, Mueket & Kates 1995; Zhang, Anninos & Norman 1995; Hernquist et al. 1996; Theuns et al. 1998) and theoretical models (Bi, Boerner & Chu 1992; Schaye 2001) have painted a picture in which low column density H I absorption lines trace the filaments of the ‘cosmic web’ and high column density absorption lines trace the surroundings of galaxies. Simulations that include self-shielding of the UVB reproduce the observed column density distribution over 10 orders of magnitude (Altaï et al. 2011).

In this paradigm, the IGM as probed by the Lyman  $\alpha$  forest consists of mildly non-linear gas density fluctuations. The gas traces the dark matter, and is photoionized and photoheated by the UVB. Although metals are detected in the IGM (Cowie et al. 1995), even at low densities (e.g. Schaye et al. 2003; Aracil et al. 2004), stirring of the IGM due to feedback from galaxies or active galactic nuclei is probably not strongly affecting the vast majority of the baryons (e.g. Theuns et al. 2002b; McDonald et al. 2005). This makes it possible to use Lyman  $\alpha$  observations to constrain cosmological parameters (McDonald & Miralda-Escudé 1999; Rollinde et al. 2003; McDonald et al. 2006; Viel & Haehnelt 2006), as well as to probe the density distribution around quasars and galaxies (Rollinde et al. 2005; Guimarães et al. 2007; Kim & Croft 2008).

Photoheating of the low-density IGM introduces a near power-law relation between its temperature,  $T$ , and density,  $\rho$ , of the form  $T = T_0 \Delta^{\gamma-1}$ , where  $\Delta \equiv \rho/\langle\rho\rangle$  (Hui & Gnedin 1997; Theuns et al. 1998). The evolution of  $T_0$  and  $\gamma$  have been measured (Ricotti, Gnedin & Shull 2000; Schaye et al. 2000; McDonald et al. 2001; Lidz et al. 2006, 2010; Becker et al. 2007, 2011), and depend on the reionization history (e.g. Theuns et al. 2002a; Hui & Haiman 2003) and the hardness of the UVB. When the gas is strongly photoheated after the reionization of H I and He II,  $T_0$  increases and the gas becomes nearly isothermal,  $\gamma \rightarrow 1$ ; asymptotically, the balance between photoheating and adiabatic cooling results in  $T = T_0 \Delta^{1/1.7}$  and a slowly decreasing  $T_0$  with redshift (Hui & Gnedin 1997; Theuns et al. 1998). The amplitude of the optically thin ionizing background rate ( $\Gamma_{12}$ ), the temperature of the IGM (characterized by  $T_0$  and  $\gamma$ ) and the amplitude of fluctuations ( $\sigma_8$ )

★ E-mail: rollinde@iap.fr

together determine the net amount of absorption (Rauch et al. 1997; Theuns et al. 2002a; Hui & Haiman 2003; Bolton et al. 2005; Fan et al. 2006; Faucher-Giguère et al. 2008), and the value inferred by comparing to simulations is very close to that computed by summing over sources by Haardt & Madau (2001).

It is also possible to compare the full probability distribution function of the transmission (TPDF) between simulations and data, which could provide a more accurate characterization of the UVB. Such an analysis was performed by Bolton et al. (2008) and Viel, Bolton & Haehnelt (2009), who compared TPDFs computed from simulations to those measured from a large sample of high-resolution UVES spectra (Kim et al. 2007). They performed a standard  $\chi^2$  analysis and suggested that an ‘inverted’  $T$ – $\rho$  relation,  $\gamma < 1$ , may be required to fit the data. A similar conclusion was reached by Becker et al. (2007) using Keck data and different theoretical optical depth distributions. Calura et al. (2012) have done the same analysis with additional quasars at  $z \gtrsim 3$ . Their new analysis favours a value of  $\gamma$  that is larger than what they found before, but is still slightly lower than 1. From a theoretical point of view, it is difficult to understand how an inverted temperature–density relation might arise: simulations that include spectral hardening computed with a full radiative transfer calculation (e.g. Bolton, Oh & Furlanetto 2009; McQuinn et al. 2009) do not result in  $\gamma < 1$ . If the IGM’s  $T$ – $\rho$  relation were indeed inverted, there may be missing physics in simulations of the Lyman  $\alpha$  forest [such as the impact of blazars as studied recently by Chang, Broderick & Pfrommer (2012) and Puchwein et al. (2014)], which may impact other statistics such as the Lyman  $\alpha$  power spectrum (e.g. McQuinn et al. 2011) and cosmological constraints derived from that (e.g. Gratton, Lewis & Efstathiou 2008; Boyarsky, Ruchayskiy & Iakubovskiy 2009). Partly for this reason, Lyman  $\alpha$  forest constraints were not used by Komatsu et al. (2009) in their determination of cosmological parameters from *WMAP* and other data.

However, there are both numerical and observational difficulties in the characterization of the absorption. Numerical issues were investigated in a paper by Tytler et al. (2009), who analysed the importance of large-scale modes in the determination of the TPDF in a numerical simulation. These authors showed that smaller simulation boxes predict, on average, more absorption for a given value of the imposed ionizing background. The box size used in the analyses of Bolton et al. (2008) is 56 Mpc, which, according to Tytler et al. (2009, their table 12), decreases the amplitude of the TPDF by 1–5 per cent in the flux range used in the analysis (0.2–0.8) as compared to a bigger box of 76.8 Mpc. The difference could be up to 10 per cent for even larger simulations. Even so, Tytler et al. (2009) also found that the predicted TPDF (with their box size of 76.8 Mpc) differs from the observed one, although to a lesser extent than that seen by Bolton et al. (2008). They did not consider an inverted temperature–density relation, but discussed other plausible sources for the discrepancy: the lack of high column density lines [ $\log_{10} N_{\text{H I}}(\text{cm}^{-2}) > 14$ ] in the simulation, unidentified metal lines and the assumed mean flux values. Note that the last two issues were discussed and, at least partly, accounted for in Bolton et al. (2008).

However, an additional limitation, not considered in Tytler et al. (2009), is the relatively small number of observed high-resolution spectra. For example, Kim et al. (2007) use a sample of just 18 spectra. In this paper, we use both simulations and data to get a better handle on just how well such a relatively small sample of spectra determines the TPDF.

We revisit the analysis of the transmission statistics in terms of its sample variance using four different observational determinations

described in Section 2.1: (i) the LUQAS sample of Kim et al. (2007) used by Bolton et al. (2008), (ii) the sample of Calura et al. (2012) that increases the number of quasar with  $z \gtrsim 3$ , (iii) a sample of Keck spectra analysed and published by McDonald et al. (2000) and (iv) finally a UVES sample collected in the context of the European Southern Observatory (ESO) Large Programme (LP) ‘Cosmic Evolution of the IGM’ (Bergeron et al. 2004). We demonstrate that published errors on the mean transmission are often too small and they do not fully account for sample variance. The observed TPDFs are compared to mock spectra computed from a suite of hydrodynamical simulations called Galaxies-Intergalactic Medium Interaction Calculation (GIMIC; Section 2.3; Crain et al. 2009) that resolves both large and small scales by using ‘zoomed’ initial conditions. We generate many mock samples from GIMIC with the same redshift path as the observed samples, and use this to investigate sample variance in both the mean transmission and the transmission probability distribution. In particular, we show how strong lines, which are relatively rare, nevertheless have substantial impact on both the mean transmission and its probability distribution, something which Viel et al. (2004) commented on in the context of the transmission power spectrum. Given the small redshift paths of the data, we conclude that observations and simulations are mutually consistent, because of the relatively ‘large sample variance’.

## 2 OBSERVED AND SIMULATED LYMAN $\alpha$ SPECTRA

### 2.1 Observed samples

The transmission in the Lyman  $\alpha$  forest is the ratio  $F = F_0/C$  of the measured flux ( $F_0$ ) over what the flux would be in the absence of absorption. Measuring  $F$  requires knowledge of the intrinsic flux of the quasar ( $C$ , the ‘continuum’), and since we are only interested in absorption due to neutral hydrogen ( $\text{H I}$  Lyman  $\alpha$ ,  $n = 1 \rightarrow 2$ ,  $\lambda_0 = 1215.57 \text{ \AA}$ ), we also need to know the contribution to the absorption from other elements (‘metals’). Neither the continuum nor the contribution from metals is easy to determine: the intrinsic QSO spectrum contains broad emission lines and, moreover, the combination of a narrow slit with an echelle spectrograph – required to obtain the high spectral resolution – means the spectra cannot be accurately flux calibrated. ‘Continuum fitting’ spectra to determine  $C$  then involves drawing a smooth curve connecting regions deemed free from absorption, a somewhat subjective procedure. Metal lines are eliminated by identifying lines too narrow to be due to hydrogen, or from line coincidences where a metal transition occurs at the same redshift as a (strong)  $\text{H I}$  absorber or other metal transition. Finally, a ‘proximity region’, i.e. the region close to the quasar where it dominates the UVB, is excised.

Here we use four observational data sets to determine the mean transmission and its PDF, referred to below as the LP sample, the LUQAS sample, the sample of Calura et al. (2012) and the McDonald et al. (2000) sample (hereafter M00 sample).

(i) The LP sample is from our own independent analysis of a set of 18 UVES VLT spectra, collected as part of the ESO’s ‘LP’ ‘Cosmic Evolution of the IGM’ (Bergeron et al. 2004). These LP spectra have a high resolution ( $\lambda/\Delta\lambda \sim 45\,000$ ) and a high S/N ( $\approx 25$ – $30$  per pixel), and were rebinned on to  $0.05 \text{ \AA}$  pixels. The continuum was fitted using an automatic method described in Aracil et al. (2004), and metal lines were removed by eliminating contaminated regions. There are no damped Lyman  $\alpha$  absorbers (DLAs) in these lines of sight. We compute the TPDFs and the mean transmission

over three relatively small redshift ranges, centred at  $z \simeq 2$  ( $1.88 < z < 2.37$ ),  $z \simeq 2.5$  ( $2.37 < z < 2.71$ ) and  $z \simeq 3$  ( $2.71 < z < 3.21$ ). The total number of data pixels in the LP spectra for each of the redshift bins is 139 830, 65 067 and 30 800 (of which a fraction of 74, 85 and 100 per cent are in common with the LUQAS sample described below). The corresponding absorption distances<sup>1</sup> are  $\Delta X = 10.5$ , 5.8 and 2.9, respectively.

(ii) The LUQAS sample used by Bolton et al. (2008) and Viel et al. (2009) is described in detail by Kim et al. (2007), including details of their method of continuum fitting and metal line identification. They fit metal lines in the Lyman  $\alpha$  part of the spectrum using VPFIT (Carswell et al. 1987) and then use this to reconstruct an H I spectrum without the identified metals, as in Theuns et al. (2002c). We find that this method has a similar effect on the transmission distribution as the method we used. The LUQAS sample has 18 spectra, 14 of which are part of the LP sample. Pixels within the Lyman  $\alpha$  forest within a given redshift range are extracted and combined into a histogram. We will refer to these published values as the ‘LUQAS’ data. The TPDFs of Kim et al. (2007) are averaged over the same redshift ranges as the LP ones.

(iii) The Calura et al. (2012) sample is used to investigate the TPDF at redshift  $z \gtrsim 3$ . Their results are split in two bins,  $2.62 < z < 3.17$  and  $3.17 < z < 3.72$ . We consider the first bin only to be compared to the other determinations. The absorption distance in this bin, after removal of 14 DLA and Lyman Limit System (LLS) regions, is about 4.5. We use their estimate of the TPDF without metals and LLS.

(iv) The M00 sample is a set of eight Keck HIRES spectra with resolution and S/N similar to the UVES data, and is described in McDonald et al. (2000). They use slightly different redshift bins that do not cover our lowest redshift bin, and go up to  $z = 4.43$ . We will therefore only consider their two lower redshift bins:  $2.09 < z < 2.67$  (33 791 data pixels,  $\Delta X \simeq 3.5$ ) and  $2.67 < z < 3.39$  (31 897 data pixels,  $\Delta X \simeq 3.7$ ).

Noise and errors in the continuum fitting can make the transmission  $F < 0$  or  $F > 1$ . To compute the PDF of the transmission for the LP sample, we use the same binning as used in the LUQAS and McDonald et al. (2000) analyses, i.e. bins of width 0.05 between  $F = 0.025$  and 0.975, plus extra bins for those pixels with  $F < 0.025$  and  $F > 0.975$ . The PDF is then normalized<sup>2</sup> such that the sum of all values in all bins equals 20. The full covariance matrix of errors on the PDF is estimated using the jackknife technique described in Lidz et al. (2006), but applied to the flux, while they applied this technique to  $\delta_f \equiv (F - \bar{F})/F$ . Specifically, we estimate the PDF  $P(F_i)$  from the full data sample, divide the data set into 30 different subgroups and then estimate the PDF of the data sample omitting each subgroup iteratively,  $P_k(i)$ . The variance  $\sigma_{i,j}$  is then computed on the difference between  $P(F_i)$  and  $P_k(F_j)$ :  $\sigma_{i,j}^2 = \sum_{k=1}^{30} [P(F_i) - P_k(F_i)][P(F_j) - P_k(F_j)]$ . For the other observations, we use error bars taken from the corresponding references. We discuss below how errors can be more reliably estimated as the variance among mock GIMIC samples. Both estimates of errors are shown in Fig. 3, while Table 1 indicates the variance among mock GIMIC samples.

<sup>1</sup> The absorption distance  $dX/dz \equiv (1+z)^2 [\Omega_m(1+z)^3 + \Omega_\Lambda]^{-1/2}$ , and quoted numerical values of  $dX$  assume  $\Omega_m = 0.25$  and  $\Omega_\Lambda = 0.75$ .

<sup>2</sup> Pixels with  $F < 0$  and  $F > 1$  are assigned to the first and last PDF bins, respectively, but the number of values in each bin is divided by the same  $\Delta F = 0.05$  bin width when normalizing the histogram.

**Table 1.** The mean TPDF of 18 UVES LP QSOs, in three redshift bins ( $1.88 < z < 2.37$ ,  $2.37 < z < 2.71$  and  $2.71 < z < 3.21$ ). The error is the  $2\sigma$  variance among mock GIMIC samples with ensemble average mean transmission  $\langle F \rangle = 0.86$ , 0.77 and 0.71, respectively.

$F$ bin centre	$\langle z \rangle = 2.0$	PDF and its error $\langle z \rangle = 2.5$	$\langle z \rangle = 3.0$
0.00	$0.6052 \pm 0.0990$	$1.2092 \pm 0.1840$	$1.6649 \pm 0.4680$
0.05	$0.2004 \pm 0.0390$	$0.4044 \pm 0.0670$	$0.4466 \pm 0.1520$
0.10	$0.1472 \pm 0.0240$	$0.2734 \pm 0.0390$	$0.3130 \pm 0.0850$
0.15	$0.1471 \pm 0.0220$	$0.2211 \pm 0.0300$	$0.2894 \pm 0.0700$
0.20	$0.1380 \pm 0.0220$	$0.1823 \pm 0.0320$	$0.2441 \pm 0.0690$
0.25	$0.1370 \pm 0.0210$	$0.2253 \pm 0.0290$	$0.2690 \pm 0.0680$
0.30	$0.1383 \pm 0.0220$	$0.2228 \pm 0.0300$	$0.2468 \pm 0.0660$
0.35	$0.1350 \pm 0.0230$	$0.2062 \pm 0.0310$	$0.2527 \pm 0.0620$
0.40	$0.1539 \pm 0.0240$	$0.2291 \pm 0.0310$	$0.2423 \pm 0.0660$
0.45	$0.1602 \pm 0.0260$	$0.2797 \pm 0.0350$	$0.2568 \pm 0.0670$
0.50	$0.1815 \pm 0.0270$	$0.2780 \pm 0.0340$	$0.2745 \pm 0.0670$
0.55	$0.2029 \pm 0.0280$	$0.2877 \pm 0.0340$	$0.3474 \pm 0.0750$
0.60	$0.2253 \pm 0.0290$	$0.3514 \pm 0.0360$	$0.4180 \pm 0.0810$
0.65	$0.2855 \pm 0.0320$	$0.3899 \pm 0.0380$	$0.5014 \pm 0.0930$
0.70	$0.3341 \pm 0.0370$	$0.4519 \pm 0.0420$	$0.5879 \pm 0.0980$
0.75	$0.4120 \pm 0.0410$	$0.5815 \pm 0.0520$	$0.7192 \pm 0.1210$
0.80	$0.5508 \pm 0.0480$	$0.8224 \pm 0.0650$	$0.8886 \pm 0.1390$
0.85	$0.8279 \pm 0.0610$	$1.1295 \pm 0.0850$	$1.3261 \pm 0.1840$
0.90	$1.3857 \pm 0.0810$	$1.7072 \pm 0.1070$	$1.8837 \pm 0.2460$
0.95	$3.5231 \pm 0.1240$	$3.1205 \pm 0.1760$	$2.9413 \pm 0.4360$
1.00	$10.1090 \pm 0.4230$	$7.4264 \pm 0.4510$	$5.8861 \pm 0.8010$

## 2.2 Inconsistency between measured values of the mean transmission

We compare estimates of the mean transmission collected from the literature (McDonald et al. 2000; Kirkman et al. 2005; Kim et al. 2007; Faucher-Giguère et al. 2008), as well as measured by us for the LP sample. Errors are based on a bootstrap procedure, by resampling chunks of spectra of size  $5 \text{ \AA}$ , or on the variance among chunks of the same size (Faucher-Giguère et al. 2008, hereafter FG). Kim et al. (2007) only provide errors on the effective optical depth, for a smaller bin in redshift  $dz = 0.2$ . We quote the corresponding errors on the flux  $\sigma_F = F \sigma_\tau$ , and we compute bootstrap errors for the LP using the same bins in redshift. Estimates from LP and LUQAS are given in Table 2 (upper rows), with corresponding  $2\sigma$  errors, scaled to the same absorption distance.

The mean transmission values obtained from the LUQAS and LP samples differ by  $2.13\sigma$ ,  $2.40\sigma$  and  $2.75\sigma$  at  $z = 2$ , 2.5 and 3, respectively (where  $\sigma$  is obtained from adding the bootstrap errors from both samples in quadrature). We recall that the LUQAS and LP samples are mostly based on the *same* raw data, but that those data were reduced by different groups. These differences must therefore be due to systematic errors in the adopted procedures, in particular differences in continuum fitting and the treatment of absorption from metals. Also, Kim et al. (2007) concluded that the treatment of the data, in particular continuum fitting, leads to notable differences between authors. Published values for  $\bar{F}$  from LUQAS, Kirkman et al. (2005) and FG agree within  $1\sigma$  at  $z = 2$ , but the differences increase at higher  $z$ . The most discrepant values are  $2.49\sigma$  at  $z = 2.5$  [LUQAS versus McDonald et al. (2000)], both are high-resolution data] and  $3.9\sigma$  at  $z = 3$  [Kirkman et al. (2005) versus FG].

How reliable are the quoted errors? Kim et al. (2007) estimate errors on the effective optical depth,  $-\ln(\bar{F})$ , by bootstrapping the LUQAS spectra in chunks of  $5 \text{ \AA}$ . They do not mention convergence

**Table 2.** Upper rows: measured value of the mean transmission in three redshift ranges for LUQAS and LP samples. Using the LP as a reference, the redshifts ranges are 1.88–2.37, 2.37–2.71 and 2.71–3.21 with absorption distance of 10.3, 5.8 and 2.9, respectively. For LUQAS, Kim et al. (2007) provide errors computed by bootstrapping chunks of size  $5 \text{ \AA}$  within bins of size  $\Delta z = 0.2$ ; their errors are then rescaled to the LP absorption distances. The errors given for LP correspond to the variance between GIMIC mock samples. Lower rows: ensemble-averaged  $\langle F \rangle$  in GIMIC simulations that reproduce within  $2\sigma$  the LP observed TPDF and mean transmission  $\bar{F}$ ; the last row gives the ionizing background rate values in the same GIMIC simulations.  $\langle F \rangle$  refers to an ensemble average,  $\bar{F}$  refers to a single realization of such an ensemble and is generally larger than  $\langle F \rangle$  because it includes a 2 per cent continuum fitting offset.

$\bar{z} = 2.0$	$\bar{z} = 2.5$	$\bar{z} = 3.0$	
Measured $\bar{F} (\pm 2\sigma)$			
$0.887 \pm 0.011$	$0.812 \pm 0.017$	$0.780 \pm 0.034$	LP
$0.868 \pm 0.010$	$0.775 \pm 0.021$	$0.713 \pm 0.032$	LUQAS
Derived $\langle F \rangle$ with $2\sigma$ variance from GIMIC mocks			
$0.86^{+0.007}_{-0.025}$	$0.77^{+0.005}_{-0.045}$	$0.71^{+0.06}_{-0.05}$	(from TPDF)
$0.85 \pm 0.02$	$0.79^{+0.02}_{-0.04}$	$0.71^{+0.07}_{-0.09}$	(from $\bar{F}$ )
Derived $\Gamma_{12}$ with $2\sigma$ range from GIMIC mocks			
1.3 (0.9, 2.0)	1.2 (0.8, 1.8)	1.3 (0.6, 2.6)	

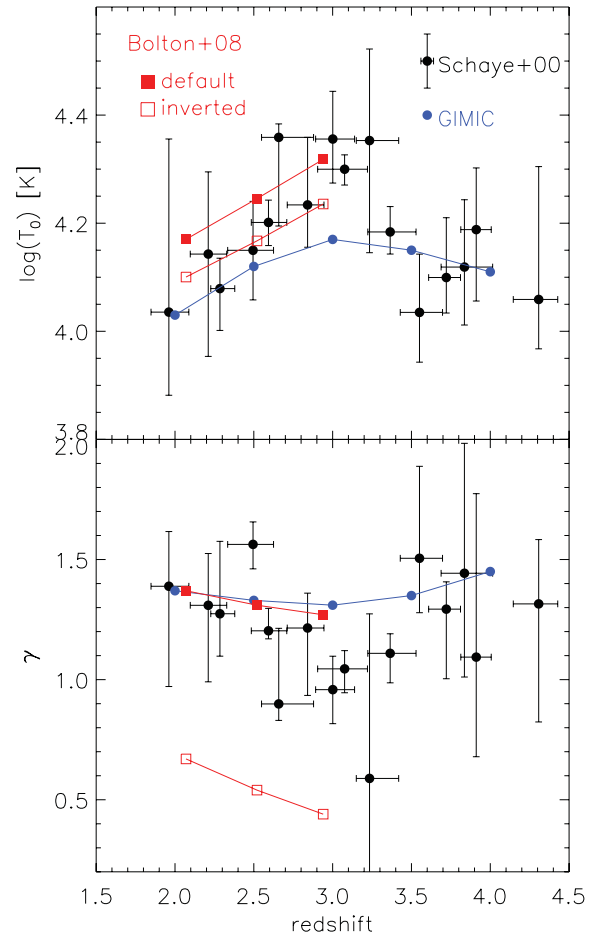
tests with chunk size for the error on the mean flux, but they do note that a modified jackknife method, using  $50 \text{ \AA}$  chunks, yields errors that are too low – comparable to the estimated variance due to continuum placement alone. They nevertheless use jackknife errors with  $50 \text{ \AA}$  chunks to compute the variance of the TPDF. Calura et al. (2012) compare errors on the TPDF estimated with a bootstrap on  $5 \text{ \AA}$  chunks and with a jackknife on  $50 \text{ \AA}$  chunks. They find similar results, but do not mention convergence tests with chunk size either. FG mention that ‘We have verified that the error estimates have converged for our choice of segment length’, but they do not present quantitative results.

Bootstrap errors depend on the arbitrary size of the chunks from which they are computed. Indeed, for the LP data at  $z = 2.5$ , we find variances in the mean flux of  $\sigma = [0.25, 0.53, 0.78, 1.14, 1.03, 1.33, 1.15, 1.16] \times 10^{-2}$  for chunk sizes of  $[0.2, 1, 5, 25, 50, 125, 250, 625] \text{ \AA}$ . Although  $\sigma$  converges for very large chunk sizes  $\gtrsim 25 \text{ \AA}$ , as expected, we suggest that typical published errors based on  $5 \text{ \AA}$  chunks underestimate the variance by  $\sim 50$  per cent. Note that the largest chunk size we tested,  $625 \text{ \AA}$ , is comparable to the extent of the Lyman  $\alpha$  forest in a  $z \sim 3$  QSO. We discuss the reliability of bootstrap errors using GIMIC mocks further in Section 2.4 below.

### 2.3 Mock samples

We use the GIMIC (Crain et al. 2009) simulations, a set of smoothed particle hydrodynamics simulations of five nearly spherical regions of comoving radius  $R \sim 18 h^{-1} \text{ Mpc}$  picked from the Millennium Simulation (Springel et al. 2005). The simulations have a gas particle mass of  $1.4 \times 10^6 h^{-1} M_{\odot}$ . These ‘zoomed’ simulations allow us to obtain high numerical resolution and yet include the effects of large-scale power, i.e. the simulation probes a range of environments, from massive clusters to deep voids. The effect of large-scale structures, as discussed in Tytler et al. (2009), is thus accounted for.

The GIMIC simulations were performed with the GADGET-3 code, an evolution of GADGET-2 described last by Springel (2005), with modules for star formation, feedback from galactic winds, chemodynamics and radiative cooling and photoheating due to an imposed evolving UVB, as described in Schaye & Dalla Vecchia (2008), Dalla Vecchia & Schaye (2008), Wiersma, Schaye & Smith (2009a) and Wiersma et al. (2009b), respectively (see also Schaye et al. 2010). The assumed cosmological parameters are  $(\Omega_{\text{cdm}} + \Omega_{\text{b}}, \Omega_{\Lambda}, \Omega_{\text{b}}, n_{\text{s}}, h, \sigma_8) = (0.25, 0.75, 0.045, 0.9, 0.73, 0.9)$ . The five GIMIC regions are picked such that their overdensities at redshift  $z = 1.5$  are  $(-2, -1, 0, 1, 2)$  times the rms deviation,  $\sigma$ , from the mean on the spatial scale of the spheres. Reionization of H I is assumed to occur at  $z = 9$ , heating the IGM to  $T \sim 10^4 \text{ K}$ , and of He II at  $z = 3.5$ . As also shown by Wiersma et al. (2009b), the evolution of  $T_0$  and  $\gamma$  in the simulations are broadly consistent with the Schaye et al. (2000) measurements (see also Fig. 1). For densities close to the mean,  $\gamma \gtrsim 1.3$  and the temperature–density relation is never ‘inverted’.



**Figure 1.** Evolution of the parameters  $T_0$  and  $\gamma$  of the temperature–density relation  $T = T_0 (\rho/\langle\rho\rangle)^{\gamma-1}$ , as measured by Schaye et al. (2000, black circles with error bars) and in the GIMIC simulation (blue connected dots). The temperature–density relation in GIMIC is broadly consistent with the measured values. He II reionization causes the rise in  $T_0$  and the corresponding dip in  $\gamma$  in the GIMIC simulations at redshift  $z \sim 3.2$ , but  $\gamma$  never drops below  $\sim 1.3$ . Red symbols are from the model of Bolton et al. (2008): filled squares are for their default model, open squares are for their model 20-256g3 that best fits the TPDF they inferred from LUQAS. This model has an inverted temperature–density relation, i.e.  $\gamma < 1$ .



We compute 1000 mock Lyman  $\alpha$  forest spectra by tracing straight lines through a cube<sup>3</sup> embedded well within each of the five spheres, extracting density, temperature and peculiar velocity along them, and then computing the corresponding optical depth as described in Theuns et al. (1998). Crain et al. (2009) explain in their appendix how to combine results from individual spheres to correctly reproduce statistics valid for the full Millennium volume: we use the weights listed in their table A1. Given these weights, we generate a ‘mock’ LP sample by randomly selecting spectra from each of the five spheres until the redshift path of mock and LP samples are the same. We repeat this procedure 400 times to obtain a ‘suite’ of mock samples. Note that every single mock sample in the suite has the same redshift path as the LP sample. Each spectrum is convolved with a Gaussian to match the UVES spectral resolution, rebinned to the UVES pixel size and we add noise with similar statistical properties as measured in the observed spectra. Our results do not change significantly if we only use the GIMIC mean density sphere. We can compute flux statistics for a given mock sample simply from all pixels in all short spectra that make up the mock sample. However, when computing bootstrap errors below, we combine these short spectra into a Lyman  $\alpha$  spectrum that mimics the full absorption distance of a given LP spectrum.

It is difficult to accurately mimic the effect of ‘continuum fitting’ as applied to observations to the simulated samples, because the wavelength range over which the observed continuum is supposed to vary is large compared to the size of an individual simulated spectrum. In the observations, the true and estimated continua are thought to differ by about 1–3 per cent (see e.g. Aracil et al. 2004; FG). Therefore, to investigate plausible continuum uncertainties, we compare statistics from the original samples to those in which we multiply the flux by a constant factor of 1.02 to mimic a 2 per cent systematic offset between ‘true’ and ‘fitted’ continua.

The Lyman  $\alpha$  optical depth in a spectrum depends on the evolving photoionization rate,

$$\Gamma = 4\pi \int_{\nu_T}^{\infty} \frac{J(\nu)}{h\nu} \sigma_{\nu} d\nu \equiv \Gamma_{12} 10^{-12} \text{ s}^{-1}, \quad (1)$$

where  $J(\nu)$  is the mean intensity of the ionizing radiation at a given redshift,  $\nu_T$  is the frequency of the Lyman limit and  $\sigma_{\nu}$  is the hydrogen photoionization cross-section. Within a suite of mock samples, we use the same value for  $\Gamma_{12}$ , and will refer to the ‘ensemble average’ mean transmission of the suite as  $\langle F \rangle$ . The mean transmission,  $\bar{F}$ , of a given mock sample can differ significantly from the ensemble average  $\langle F \rangle$  of the corresponding suite because of ‘sample variance’, and the same is true for its PDF. We estimate the sample variance in a given suite by comparing all 400 mock samples that make up the suite. We emphasize that, because the simulated samples keep probing the same density field, the real dispersion is likely to be larger than this estimate.

The value of the photoionization rate  $\Gamma_{12}$  is uncertain. Theuns et al. (1998) show that, in the optically thin case, simulations can be run with one value for  $\Gamma_{12}$  and later accurately scaled to another value. To investigate the effect of uncertainties in  $\Gamma_{12}$ , we generate many suites of mock samples, with different values of  $\Gamma_{12}$  and hence of the ensemble average transmission,  $\langle F \rangle$ .

## 2.4 Estimates of errors with mock samples

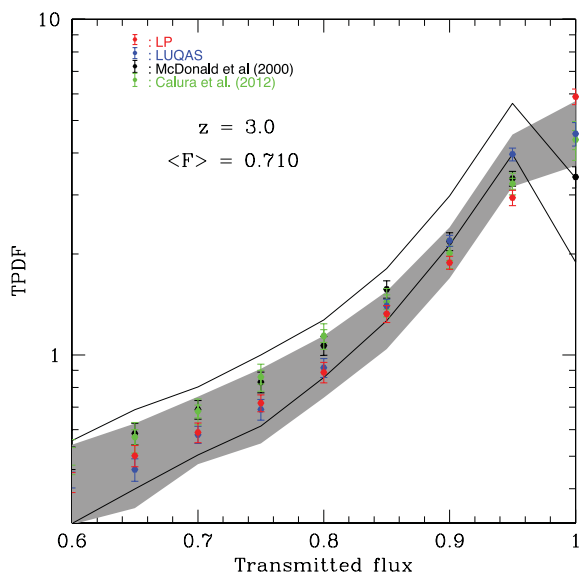
We can check the reliability of the bootstrap errors discussed in Section 2.2 using GIMIC mock samples. We first examine whether mocks generated from the simulation give the same errors on the mean flux as observed samples when the errors are estimated in the same way. FG divide the variance  $\sigma_i$  of the mean flux measured along chunks of 3 Mpc proper size by the square root of the number of chunks. They find  $\sigma_i = [0.13, 0.11, 0.09]$  at  $z = [3, 2.4, 2]$ , with 193, 263 and 50 chunks, respectively. Applying this procedure first to the LP data, we find  $\sigma_i = [0.125, 0.13, 0.095]$  at  $z = [3, 2.5, 2]$ , with 37, 262 and 413 chunks, respectively. Applied to our mocks, we find  $\sigma_i = [0.14, 0.14, 0.11]$ . Therefore, both our analysis of the LP observations and of the GIMIC simulations give error estimates in reasonable agreement with those obtained by FG. Kim et al. (2007) estimate errors on the effective optical depth,  $-\ln(\bar{F})$ , by bootstrapping the LUQAS spectra in chunks of size 5 Å. We concentrate on their estimate at  $\bar{z} = 2.59$  with a bin in redshift of  $\Delta z = 0.2$ , corresponding to a velocity path of  $88\,682 \text{ km s}^{-1}$ . We use the GIMIC simulations to generate many mock versions of the LUQAS sample, each with the same velocity path, and estimate the variance  $\sigma$  for the same chunk size. The average value for our mocks is  $\sigma_F = F\sigma_{\tau} = 0.0124$ , identical to their bootstrap error. Finally, we compare errors estimated from GIMIC against our own bootstrap errors obtained from the LP data, as discussed in the previous section. At  $z = 2.5$  and for a velocity path of  $\sim 190\,000 \text{ km s}^{-1}$ , we calculate bootstrap variances of  $\sigma = [0.26, 0.54, 0.80, 0.98, 1.22, 1.15] \times 10^{-2}$  for chunk sizes of  $[0.2, 1, 5, 25, 125, 625] \text{ Å}$  for the simulated mocks, as compared to  $\sigma = [0.25, 0.53, 0.78, 1.14, 1.33, 1.16] \times 10^{-2}$  for the LP observational data. We conclude that errors computed from GIMIC mocks are in excellent agreement with published errors, as well as errors obtained by us from the LP data, when simulated and observed errors are calculated in the same way.

The bootstrap errors discussed above clearly depend on the value of the chunk size for which they are computed, both for the data and for the simulated spectra. They start to converge for relatively large chunk sizes of  $\sim 25 \text{ Å}$ , although the convergence is not yet clearly reached. Using simulations we can also calculate the variance between different mock samples: simply generate many mock samples for a given simulation, each with the same redshift path as a given observed sample, and evaluate the variance between mock samples. This variance is  $[0.55, 0.88, 1.7] \times 10^{-2}$  at redshifts  $z = [2, 2.5, 3]$ , as compared to bootstrap errors using 25 Å chunks of  $[0.50, 0.98, 2.1] \times 10^{-2}$ , in reasonable agreement. Given the dependence of the variance on chunk size for small chunks, we will use the variance between mock samples to characterize the expected level of scatter in the data and to investigate the consistency between simulation and data. We suggest that error estimates that we obtain from determining the variance between mocks are more realistic than the published, observed bootstrap errors.

## 3 THE TRANSMISSION PDF

We have computed the TPDFs of the LP sample over the same small redshift ranges as used by Kim et al. (2007). Because these redshift ranges are relatively narrow, evolution over them can be safely neglected, and hence we simply use simulation snapshots at a single redshift ( $z \simeq 2, 2.5$  and 3 for the three bins used by Kim et al. 2007) when comparing to the observed data.

<sup>3</sup> The cubes have sides  $\sim 11 h^{-1}$  comoving Mpc which ensures we stay well away from the edges of the spheres to avoid artificial boundary effects (see Crighton et al. 2010, for details). We will call a Lyman  $\alpha$  spectrum obtained from a single cut through the cube a *short* spectrum.



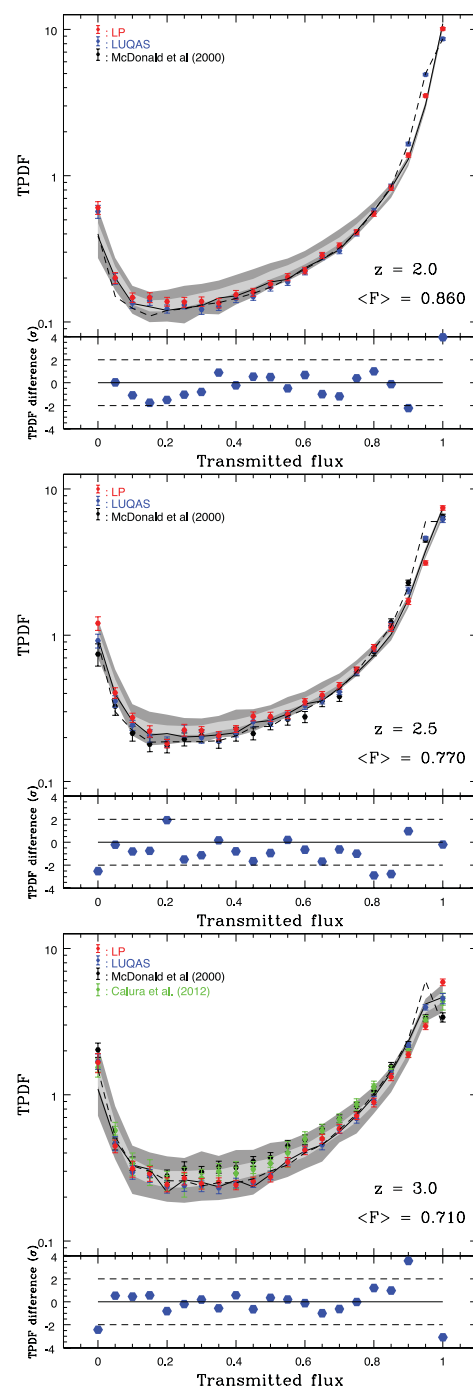
**Figure 2.** Effect of ‘continuum fitting’ the GIMIC simulations; solid curves show the  $2\sigma$  range in the TPDFs of a sample of mocks with given ensemble-averaged transmission,  $\langle F \rangle$ . When errors in continuum fitting are mimicked by a systematic shift in the continuum (see Section 2.2), the range is enclosed by full lines. Continuum fitting makes the shape of the TPDF uncertain close to  $F = 1$ . Note that we only show the range  $0.6 \leq F \leq 1$ . For  $F < 0.6$ , we find that the continuum correction is small compared to the  $2\sigma$  range. Symbols with error bars are the data from the LP sample (red), LUQAS (blue), McDonald et al. (2000) (black) and Calura et al. (2012) (green). These also show significant differences in the range  $F > 0.7$ , plausibly due to the different continuum fitting methods applied in the data reduction.

### 3.1 Variance of the transmission PDF

Fig. 2 illustrates that continuum fitting quite noticeably affects the TPDF near  $F \simeq 1$ , and comparison to the overplotted data also suggests that uncertainties in continuum placement can explain the large differences in the *observed* PDFs at  $F \simeq 1$ . Recall that we mimic the errors in continuum fitting by a systematic shift in the continuum (Section 2.2). Clearly, given these uncertainties, this part of the TPDF cannot constrain models robustly (see also Meiksin, Bryan & Machacek 2001). Fortunately, the distribution of pixels with  $F < 0.7$ , say, is relatively insensitive to the error in the continuum placement for high-resolution spectra and can thus be used to constrain the mean transmitted flux.

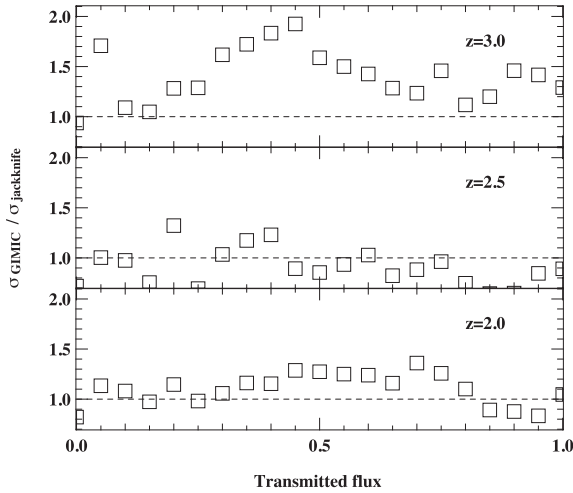
The GIMIC simulations that best reproduce the observed TPDFs for  $F < 0.7$  have ensemble-averaged mean transmissions of  $\langle F \rangle = 0.86, 0.77$  and  $0.71$  at redshifts  $z = 2, 2.5$  and  $3$ , respectively, as discussed in more detail below. Observed and mock TPDFs with these values of  $\langle F \rangle$  are compared at  $z = 2, 2.5$  and  $3$  in Fig. 3. Light (dark) shaded regions show the  $1\sigma$  ( $2\sigma$ ) dispersion<sup>4</sup> among TPDFs of this particular suite of mocks. There is considerable variance between the TPDFs of mock realizations, even though each mock realization is generated from the same simulation with the full absorption distance of the LP observed sample.

The variance in the mocks increases with redshift since the redshift path decreases. The ratio of variance computed from GIMIC mock versus jackknife variance is shown in Fig. 4. Except at  $z = 2.5$ , variance in mocks is systematically larger, from 10 to



**Figure 3.** Top to bottom: PDF of the transmission at  $z = 2, 2.5$  and  $3$ , of the best-fitting simulations [continuum fitted GIMIC simulation: solid curve; Bolton et al. (2008) model 20-256g3 shown as open squares in Fig. 1: dashed curve], compared to observational data (symbols with error bars, LP sample: red; LUQAS: blue; Calura et al.: green; M00:  $z = 2.41$  and  $3.0$ , black). Error bars are  $1\sigma$  jackknife errors for LUQAS, LP and Calura et al., and bootstrap of  $5 \text{ \AA}$  chunks for M00. Light (dark) shaded regions correspond to the  $1\sigma$  ( $2\sigma$ ) range computed from 400 mock LP samples in GIMIC simulations with redshift and ensemble-averaged mean transmission  $\langle F \rangle$  as indicated in each panel. The simulations and various data sets agree well within the  $2\sigma$  range at all three redshifts. Insets show  $(\text{model} - \text{data})/\sigma_0$ , where model is the best-fitting PDF for GIMIC, data and  $\sigma_0$  are the LP PDF and the variance estimated in GIMIC simulations. The GIMIC simulations fit the data for  $F < 0.7$  even though  $\gamma > 1$  at all  $z$ . For  $F > 0.7$  and  $z \leq 2.5$ , different data sets are inconsistent and sensitive to continuum fitting (missing points are above 4).

<sup>4</sup> They correspond to the 2.275, 15.8655, 84.13 and 97.725 percentiles computed from 400 realizations.



**Figure 4.** Ratio of PDF variance computed from GIMIC mocks to variance computed from jackknife method for bins in transmitted flux in three different bins in redshift.

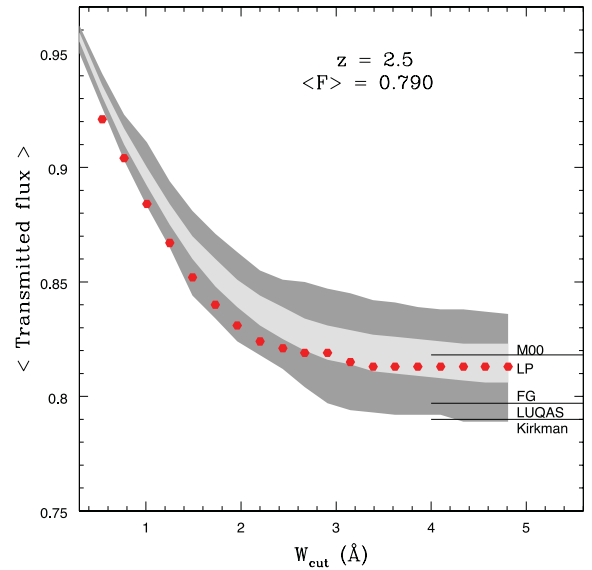
50 per cent at  $z = 2$  and up to 100 per cent at  $z = 3$ . Given that the simulations, if anything, *underestimate* sample variance, it suggests once more that the observationally determined jackknife errors are too small. Although more difficult to assess from other works, we found that the estimates of errors using the jackknife method are very unstable given the relatively small size of the sample. We will therefore quote variances computed from our mocks only.

The LP and LUQAS data fall well within the  $2\sigma$  region at all  $z$  for  $F < 0.7$ , with a possible exception of the  $F \simeq 0$  bin at  $z = 2$ . It is possible that the latter discrepancy is due to the fact that simulations that assume the gas to be optically thin do not reproduce the observed number of strong lines (e.g. Tytler et al. 2009). Including self-shielding appears to solve this issue (Alty et al. 2011). The LP and LUQAS sample results are almost identical in bins where uncertainties in the position of the continuum do not interfere in the TPDF. They are also very similar to the results from Calura et al. (2012) sample that has one quasar in common (which makes one-fourth of the total sample in this redshift bin). They also agree with results from McDonald et al. (2000) within the  $2\sigma$  range estimated from the simulations.

The difference between the best-fitting simulated PDFs in GIMIC mock samples (among different values for  $\Gamma_{12}$  only) and our determination of the TPDF from the LP, divided by  $1\sigma$  range on mock LP TPDF in GIMIC simulation, is shown in the bottom of each panel in Fig. 3. There is no evidence that the observed and simulated GIMIC PDFs are inconsistent at any redshift. The statistical interpretation of this measurement and the derived constraints on the ionizing background rate are discussed further in Section 4.

### 3.2 Variance of the mean transmission

Interestingly, observations as well as simulations show large quasar-to-quasar variations in the mean transmission at a given redshift. To illustrate the origin of this large scatter, we analyse 400 mock samples from GIMIC generated with a given ensemble average,  $\langle F \rangle = 0.79$ , at redshift  $z = 2.5$ . The large scatter is due to strong absorption lines, which contribute significantly to the mean opacity: the small number of strong lines per QSO spectrum introduces the observed scatter, as we now show (see also Desjacques, Nusser & Sheth 2007).



**Figure 5.** Mean transmission of spectra that include all lines with equivalent width  $W < W_{\text{cut}}$  for the LP sample (red dots), and the corresponding  $1\sigma$  and  $2\sigma$  range in this quantity estimated from GIMIC mock samples (grey and dark regions, respectively). The net mean transmission values,  $\bar{F}$ , for the LUQAS, M00, FG and Kirkman et al. (2005) data are indicated by horizontal lines (FG and LUQAS values of  $\bar{F}$  are identical). There is significant scatter in  $\bar{F}$  of the GIMIC samples when  $W_{\text{cut}} \gtrsim 1.5 \text{ \AA}$ , but as strong lines are excised, the dispersion decreases significantly. This shows that strong lines are mostly responsible for the scatter. The *observed* values of the net mean transmission,  $\bar{F}(W < \infty)$ , are well within the  $2\sigma$  range estimated from the GIMIC simulations.

We have used a simple criterion to identify ‘lines’ in the spectrum as regions between two maxima in  $F$ ; we also demand that the corresponding minimum is sufficiently different from the lowest maximum to avoid identifying noise features as lines. More specifically, this algorithm identifies all local minima and maxima on a spectrum smoothed with a Gaussian kernel of width  $8 \text{ km s}^{-1}$ . A line consists of all pixels between two maxima that satisfy the following two conditions: (i) two successive maxima must be separated by more than  $8 \text{ km s}^{-1}$  and (ii) the flux difference between the maxima and the minimum they straddle must be larger than four times the estimated error per pixel. Each pixel is then assigned to a line, with given equivalent width  $W$ . We can now compute the mean transmission in a mock sample (or the LP data) for all pixels in lines with  $W$  less than some maximum equivalent width,  $W_{\text{cut}}$ .

The mean transmission,  $\bar{F}(W_{\text{cut}})$ , for all pixels in lines weaker than a given value of  $W_{\text{cut}}$  is plotted as a function of  $W_{\text{cut}}$  in Fig. 5 as red dots for the LP sample, with grey and dark regions the  $1\sigma$  and  $2\sigma$  range estimated from the mock GIMIC samples. For a high cut in  $W$ , all pixels are used and  $\bar{F}(W_{\text{cut}} = \infty)$  is simply the net mean transmission  $\bar{F}$ ; we also indicate  $\bar{F}$  from LUQAS, M00, FG and Kirkman et al. (2005).

For mock samples with ensemble average  $\langle F \rangle = 0.79$ , we find that the (continuum fitted)  $\bar{F}(W_{\text{cut}} = \infty)$  varies between 0.79 and 0.84 within  $2\sigma$ . Note that our procedure to estimate the errors due to ‘continuum fitting’ makes the mean transmission  $\bar{F}$  systematically higher than  $\langle F \rangle$ . Observed determinations of the mean transmission are shown with horizontal lines in the figure. It appears that, despite the large dispersion amongst observed values, they are nevertheless consistent, because the expected sample variance, as inferred from GIMIC (and consistent with bootstrap estimates using real data for



sufficiently large chunk size), is so large. The origin of the large variance is the presence of strong lines.

#### 4 CONSTRAINTS ON THE MEAN TRANSMISSION AND THE INTENSITY OF THE IONIZING BACKGROUND

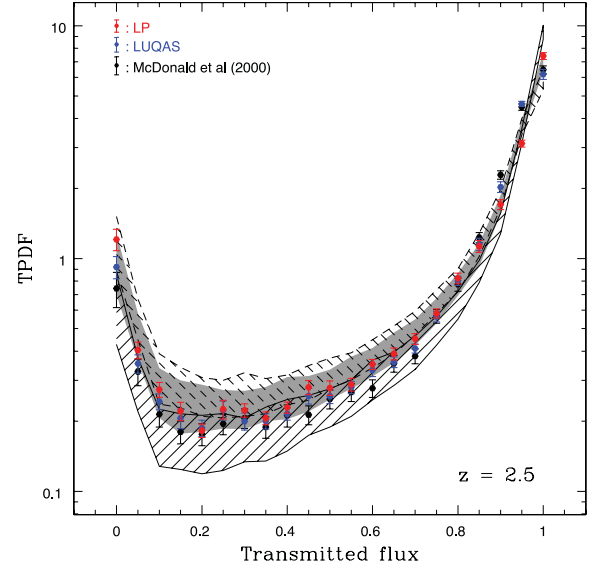
The photoionization rate can be estimated by scaling mock spectra obtained from simulations to the observed mean transmission  $\bar{F}$  and calculating the corresponding value of  $\Gamma_{12}$ . To determine the range of  $\Gamma_{12}$  values consistent with the observed  $\bar{F}$ , we need some measure of the expected variance of  $\bar{F}$  around its ensemble average  $\langle F \rangle$ . In principle, it should also be possible to use the full TPDF rather than just its mean.

To judge how well a given realization of a mock TPDF fits an observational determination, one could use the usual  $\chi^2$  estimator for values of the transmission between 0.1 and 0.7. A covariance matrix can be computed by cross-correlating estimates of the TPDF from a large number of bootstrap samples, as described in Lidz et al. (2006). Note that all bootstrap samples are then by construction subsamples of the observed spectra, which limits their usefulness if the observed path length is small. When this is applied to the TPDF, it transpires that the covariance matrix is nearly singular and hence needs to be ‘regularized’ using a singular value decomposition. We found that the values obtained for  $\chi^2$  then depend strongly on the number of singular values regularized, which severely compromises the usual statistical interpretation of  $\chi^2$ . We can get around this problem by using the simulations to estimate the variance on either  $\bar{F}$  or the TPDF, for samples with given  $\langle F \rangle$ .

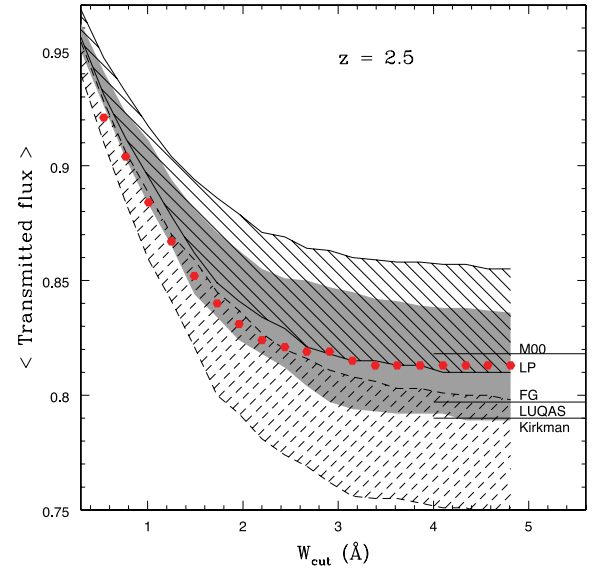
However, we have seen that the value of the mean transmission  $\bar{F}$  for a given realization of a mock sample can differ considerably from the ensemble average  $\langle F \rangle$  of the sample. Since the observations only provide a single measurement of  $\bar{F}$ , a potentially large range of ensemble averages are consistent with that  $\bar{F}$ . This is illustrated in Fig. 6 for the TPDF and in Fig. 7 for  $\bar{F}$ , both at redshift  $z = 2.5$ . In both cases, the dark grey band shows the  $2\sigma$  range in mock samples drawn from simulations with a given value of the ensemble-averaged transmission ( $\langle F \rangle = 0.77$  and  $0.79$ , respectively). As before, each sample has the same redshift path as the LP sample.

Considering first the mean transmission as a function of line width, we demand the mean transmission with  $W = \infty$  to fall within the  $2\sigma$  region. We interpret these extreme values as  $2\sigma$  limits on the ensemble average  $\langle F \rangle$ . The  $2\sigma$  allowed range is then  $0.75 \leq \langle F \rangle \leq 0.81$ . As before, the determination of  $\bar{F}$  in the mock sample is done after ‘continuum fitting’, which implies that  $\bar{F}$  will be systematically higher than  $\langle F \rangle$ . Performing the same analysis at  $z = 3$  and  $2$  yields a  $2\sigma$  allowed range of  $0.62 \leq \langle F \rangle \leq 0.78$  and  $0.83 \leq \langle F \rangle \leq 0.87$ , respectively (Table 2).

To do a fit of the TPDF requires a measure of the covariance matrix. As explained above, data samples are not yet large enough to provide a reliable estimate of it. Rather, we compute the covariance using 400 independent determinations of the TPDF in GIMIC mock samples. The covariance matrix can thus be inverted without further regularization. We use 13 bins for a range of flux  $0.1 < F < 0.7$ , corresponding to  $k = 12$  degrees of freedom. The evolution of the reduced  $\chi_r^2 = (\chi^2 - k)/\sqrt{2k}$  is shown in Fig. 8 (solid lines). To check the validity of this procedure, we derive the same evolution for different mock samples. Assuming a true value of  $\langle F \rangle_{\text{true}}$  (0.71, 0.77 and 0.86 at  $z = 2, 2.5$  and  $3$ , respectively), we compare again 400 mock samples with different value of  $\langle F \rangle$  to the average TPDF with  $\langle F \rangle_{\text{true}}$ , and compute the associated reduced  $\chi_r^2$ . The average

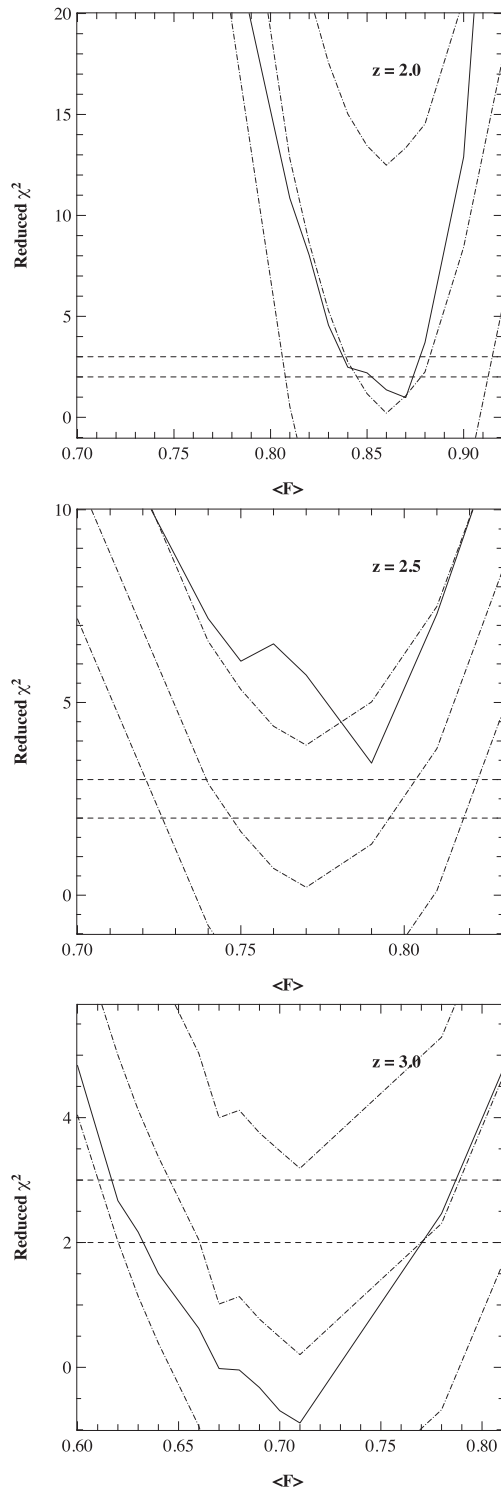


**Figure 6.** Dependence of the TPDF on the ensemble-averaged  $\langle F \rangle$  at  $z = 2.5$ . The dark shaded region shows the  $2\sigma$  range computed from 400 mock samples in a GIMIC simulation with  $\langle F \rangle = 0.77$  as in Fig. 3. Symbols with error bars are as in Fig. 3. Solid and dashed shaded regions correspond to the  $2\sigma$  range in GIMIC simulations with  $\langle F \rangle = 0.83$  and  $0.74$ , respectively. At these extremes, the observational data (for transmission  $0.1 < F < 0.7$ ) fall just outside the  $2\sigma$  range of the simulation for at least one data point.

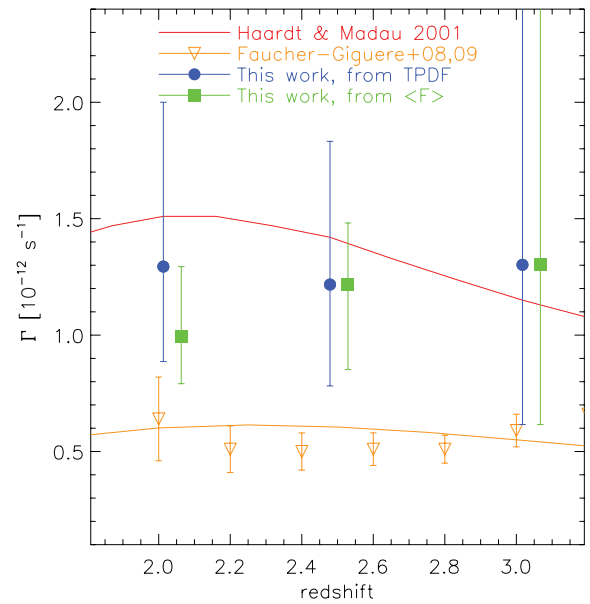


**Figure 7.** Same as Fig. 6, but for the dependence of the mean transmission as a function of maximum line width,  $\bar{F}(W_{\text{cut}})$ . Dark shaded region is the  $2\sigma$  range for  $\langle F \rangle = 0.79$ , solid and dashed shaded regions correspond to the  $2\sigma$  range in the GIMIC simulations with  $\langle F \rangle = 0.81$  and  $0.75$ , respectively.

evolution of  $\chi_r^2$  and its dispersion (dotted lines in Fig. 8) are consistent with the observed evolution using the LP TPDF, despite a slight tension at  $z = 2.5$ . We provide a best-fitting value and a  $2\sigma$  range for  $\langle F \rangle$  using the smooth average evolution in GIMIC samples:  $0.845 \leq \langle F \rangle = 0.86 \leq 0.877$  at  $z = 2.0$ ,  $0.745 \leq \langle F \rangle = 0.77 \leq 0.795$  at  $z = 2.5$  and  $0.66 \leq \langle F \rangle = 0.71 \leq 0.77$  at  $z = 3.0$ . Note that the best-fitting value for  $\langle F \rangle$  is slightly shifted compared to the value corresponding to the observed minimum, in order to best reproduce the overall evolution of  $\chi_r^2$ . Also, the range at  $z = 2.5$  as determined



**Figure 8.** Reduced  $\chi^2$  as a function of the ensemble-averaged  $\langle F \rangle$  at  $z = 2, 2.5$  and  $3.0$  (top to bottom). The covariance matrix is measured using the variance among GIMIC mock samples.  $\chi^2$  corresponds to the difference between one TPDF and the averaged TPDF from 400 GIMIC mock samples assuming different  $\langle F \rangle$ . As a validity check, the TPDF measured in one GIMIC mock sample with  $\langle F \rangle = 0.86, 0.77$  and  $0.71$  at  $z = 2, 2.5$  and  $3.0$ , respectively, is best fitted with the same value for  $\langle F \rangle$  (dotted lines show the average reduced  $\chi^2$  and the  $1\sigma$  range among 400 samples). The evolution of the reduced  $\chi^2$  as a function of  $\langle F \rangle$  is similar in the case of the observed LP (solid lines).



**Figure 9.** Mean hydrogen photoionization rate,  $\Gamma$ , as a function of redshift, from summing over sources as computed by Haardt & Madau (2001, red) and Faucher-Giguère et al. (2009, drawn orange line), and from comparing simulated to observed mock spectra. Blue and green points are our ( $2\sigma$ ) determinations from comparing, respectively, the TPDF and the mean flux in the GIMIC simulations to the LP data, orange symbols are the FG determination using a sample of 84 high-resolution quasars.

from the evolution of  $\chi^2$  is narrower than the range determined by eye in Fig. 6. Those estimates for  $\langle F \rangle$  and their  $2\sigma$  uncertainty at these three redshifts can be compared to the values given in Table 2 that refer to the allowed range of  $\langle F \rangle$  so that GIMIC simulations reproduce within  $2\sigma$  the LP observed TPDF (Fig. 6). Our values are generally in agreement with previously published values, but our quoted uncertainties are significantly larger.

Given the constraints on  $\langle F \rangle$ , we can use the simulations to infer the corresponding range in photoionization rates  $\Gamma(z)$ , which, in addition to the inferred value of  $\langle F \rangle$ , depend on the baryon density,  $\Omega_b$ , the temperature–density relation, the fluctuation amplitude  $\sigma_8$  and other cosmological parameters (Rauch et al. 1997).

Our inferred values for the photoionization rate,  $\Gamma(z)$ , are compared in Fig. 9 to the results of Haardt & Madau (2001) and to those of Faucher-Giguère et al. (2008, 2009), and are also listed in Table 2. The red (Haardt & Madau 2001) and orange (Faucher-Giguère et al. 2009) curves combine observationally inferred values for the emissivities of sources of ionizing photons with an assumed escape fraction and a model for the mean free path based on observations to estimate  $\Gamma$ . Note that Haardt & Madau (2012) derived recently a lower value of  $\Gamma \simeq 0.9 \times 10^{-12} \text{ s}^{-1}$  for  $2 < z < 3$ . In agreement with these models, we find little evidence for evolution in  $\Gamma$  over the redshift range  $z = 2$ – $3$ . This is also in agreement with the results of Bolton et al. (2005, their fig. 7), although our error bars are again larger for  $z = 2.5$  and  $3$ . Our value for the amplitude is in good agreement with that from Haardt & Madau, but is a factor of  $\sim 2$  larger than that of Faucher-Giguère et al. (2009). The latter value is not inferred from simulations, but from a fit to the density distribution of the IGM by Miralda-Escudé, Haehnelt & Rees (2000), itself guided by older simulations of Miralda-Escudé et al. (1996). The significant differences in cosmological parameters of those simulations might explain the significant offset in the inferred amplitude.

Indeed, Pawlik, Schaye & van Scherpenzeel (2009) found that the Miralda-Escudé et al. fit did not describe their own simulations well.

## 5 DISCUSSION AND CONCLUSIONS

We have compared the mean transmission,  $\bar{F}$ , as well as the TPDF in the H I Lyman  $\alpha$  forest as derived from several observational samples, as well as from mock samples computed using the GIMIC suite of hydrodynamical simulations. The mean transmission  $\bar{F}$  in the Lyman  $\alpha$  forest varies considerably from QSO to QSO, even at a given redshift. We have shown that, both in data and in simulations, this is due to the presence of strong lines, which, though relatively rare, contribute significantly to the opacity. This implies that a large redshift path is required to accurately determine the mean transmission.

We have compared in detail the variance  $\sigma$  on  $\bar{F}$  between published data, our own analysis of the observed UVES LP sample and mocks computed from the GIMIC hydrodynamical simulations. We have shown, *from observations only*, that bootstrap errors depend sensitively on chunk size, and only start to converge when relatively large chunks,  $\gtrsim 25 \text{ \AA}$ , are used. This is larger than typically used, and as a consequence we claim that published errors may be slightly underestimated, especially at larger redshift. We compared the mean transmission computed from the GIMIC simulations to that obtained from three observational samples. The GIMIC simulations are zoomed simulations of different density regions picked from the Millennium Simulation, and as such they have a realistic amount of ‘sample variance’. We exploited this feature of the simulations to estimate the uncertainty in the determination of  $\langle F \rangle$  for various observed samples. When we compute errors in the same way as performed in published work, we find excellent agreement between published and predicted values. We have also shown that converged bootstrap errors are in good agreement with errors found from bootstrapping *mock samples*. Thus, we find larger uncertainties than in previous works. For a given value of  $\langle F \rangle$ , the variance on the mean transmission is large enough to make all previously published values consistent within the scatter.

Using mock spectra derived from GIMIC, we have investigated the dependence of the variance of the mean transmitted flux on the absorption path  $\Delta X$  (see Table 3). At  $z = 2.5$ , with a sample twice as large as the LP sample, the  $2\sigma$  variance is only 0.013 and decreases down to 0.009 with a sample four times as large, which is half of the value for  $2\sigma$  for one LP sample, as expected. We note, however, that the size of our simulations may not be sufficient to evaluate the variance with such a large velocity path, especially at  $z = 2$ .

We have also investigated the probability distribution of the transmission. The ensemble variance between mock samples is systematically larger than the jackknife errors used by previous authors,

by a factor of 1.5–2 in the redshift bins  $\bar{z} = 3$ . More importantly, the covariance matrix derived from a suite of mocks can be inverted without regularization, contrary to standard estimate with jackknife methods. We used these larger errors and compare data to simulations.

The temperature–density relation,  $T = T_0 (\rho/\langle\rho\rangle)^{\gamma-1}$ , in the GIMIC simulations is a result of adiabatic cooling and photoheating due to an imposed ionizing background as computed by Haardt & Madau (2001), tweaked to yield values for  $T_0$  and  $\gamma$  consistent with the measured values of Schaye et al. (2000). In this model  $\gamma > 1$  at all times, with a minimum value of  $\gamma \simeq 1.3$  around redshift  $z = 3$  caused by He II reionization (Theuns et al. 2002a). The GIMIC TPDF is in agreement with that measured from high-resolution quasar spectra over the redshift range  $z = 2\text{--}3$  in the transmission range  $0.1 < F < 0.7$ . For  $F < 0.1$ , there may be differences due to the neglect of self-shielding in the simulations, whereas for  $F > 0.7$  uncertainties in continuum fitting the data complicate the comparison. This agreement is obtained using a specific set of cosmological parameters. In particular, we assume  $\sigma_8 = 0.9$ . The goal of this work is not to provide the best-fitting cosmological model, but to point out the large effect of sample variance. Indeed, our model with  $(\sigma_8, \gamma) = (0.9, 1)$  is not ruled out by the current set of data, while Viel et al. (2009) discard those values at more than  $2\sigma$  when considering the whole flux range. Thus, we argue that previous suggestions for an inverted  $T$ – $\rho$  relation may have resulted from an underestimate of the errors in the observations, rather than a discrepancy between data and the standard model.

## ACKNOWLEDGMENTS

We thank the anonymous referee for useful comments that improved the quality of the paper. We would like to thank our collaborators for allowing us to analyse the GIMIC simulations for this purpose. These simulations were carried out using the HPCx facility at the Edinburgh Parallel Computing Centre (EPCC) as part of the EC’s DEISA ‘Extreme Computing Initiative’, and with the Cosmology Machine at the Institute for Computational Cosmology of Durham University. This work was supported by an NWO VIDI grant and by the Marie Curie Initial training Network CosmoComp (PITN-GA-2009-238536).

## REFERENCES

- Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, *ApJ*, 737, L37
- Aracil B., Petitjean P., Pichon C., Bergeron J., 2004, *A&A*, 419, 811
- Becker G. D., Rauch M., Sargent W. L. W., 2007, *ApJ*, 662, 72
- Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, *MNRAS*, 410, 1096
- Bergeron J. et al., 2004, *The Messenger*, 118, 40
- Bi H. G., Boerner G., Chu Y., 1992, *A&A*, 266, 1
- Bolton J. S., Haehnelt M. G., Viel M., Springel V., 2005, *MNRAS*, 357, 1178
- Bolton J. S., Viel M., Kim T.-S., Haehnelt M. G., Carswell R. F., 2008, *MNRAS*, 386, 1131
- Bolton J. S., Oh S. P., Furlanetto S. R., 2009, *MNRAS*, 395, 736
- Boyarsky A., Ruchayskiy O., Iakubovskiy D., 2009, *J. Cosmol. Astropart. Phys.*, 3, 5
- Calura F., Tescari E., D’Odorico V., Viel M., Cristiani S., Kim T.-S., Bolton J. S., 2012, *MNRAS*, 422, 3019
- Carswell R. F., Webb J. K., Baldwin J. A., Atwood B., 1987, *ApJ*, 319, 709
- Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, *ApJ*, 437, L9
- Chang P., Broderick A. E., Pfrommer C., 2012, *ApJ*, 752, 23C

**Table 3.** Dependence of the variance of the mean transmission on absorption distance  $\Delta X$ , for three redshifts. The top row shows the variance ( $2\sigma$ ) for the current LP sample (with given absorption distance  $\Delta X$  LP). The second and third rows are for samples two and four times as large. Errors correspond to the variance among mock LP samples.

$\Delta X$ LP	$\bar{z} = 2.0$ 10.5	$\bar{z} = 2.5$ 5.8	$\bar{z} = 3.0$ 2.9	Sample size
	0.011	0.017	0.034	$\Delta X \times 1$
	0.0078	0.013	0.024	$\Delta X \times 2$
	0.0054	0.0088	0.017	$\Delta X \times 4$

- Cowie L. L., Songaila A., Kim T.-S., Hu E. M., 1995, *AJ*, 109, 1522
- Crain R. A. et al., 2009, *MNRAS*, 399, 1773
- Crighton N. H. M., Morris S. L., Bechtold J., Crain R. A., Jannuzi B. T., Shone A., Theuns T., 2010, *MNRAS*, 402, 1273
- Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431
- Desjacques V., Nusser A., Sheth R. K., 2007, *MNRAS*, 374, 206
- Fan X., Carilli C. L., Keating B., 2006, *ARA&A*, 44, 415
- Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, *ApJ*, 681, 831 (FG)
- Faucher-Giguère C., Lidz A., Zaldarriaga M., Hernquist L., 2009, *ApJ*, 703, 1416
- Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
- Gratton S., Lewis A., Efstathiou G., 2008, *Phys. Rev. D*, 77, 083507
- Guimarães R., Petitjean P., Rollinde E., de Carvalho R. R., Djorgovski S. G., Srianand R., Aghaee A., Castro S., 2007, *MNRAS*, 377, 657
- Gunn J. E., Peterson B. A., 1965, *ApJ*, 142, 1633
- Haardt F., Madau P., 2001, in Neumann D. M., Van J. T. T., eds, *XXIst Moriond Astrophysics Meeting. Clusters of Galaxies and the High Redshift Universe Observed in X-rays*. Savoie, France
- Haardt F., Madau P., 2012, *ApJ*, 746, 125H
- Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, *ApJ*, 457, L51
- Hu E. M., Kim T.-S., Cowie L. L., Songaila A., Rauch M., 1995, *AJ*, 110, 1526
- Hui L., Gnedin N. Y., 1997, *MNRAS*, 292, 27
- Hui L., Haiman Z., 2003, *ApJ*, 596, 9
- Kim Y.-R., Croft R. A. C., 2008, *MNRAS*, 387, 377
- Kim T.-S., Bolton J. S., Viel M., Haehnelt M. G., Carswell R. F., 2007, *MNRAS*, 382, 1657
- Kirkman D. et al., 2005, *MNRAS*, 360, 1373
- Komatsu E. et al., 2009, *ApJS*, 180, 330
- Lidz A., Heitmann K., Hui L., Habib S., Rauch M., Sargent W. L. W., 2006, *ApJ*, 638, 27
- Lidz A., Faucher-Giguère C.-A., Dall'Aglio A., McQuinn M., Fechner C., Zaldarriaga M., Hernquist L., Dutta S., 2010, *ApJ*, 718, 199
- Lynds R., 1971, *ApJ*, 164, L73
- McDonald P., Miralda-Escudé J., 1999, *ApJ*, 518, 24
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W., Barlow T., Cen R., Ostriker J., 2000, *ApJ*, 543, 1
- McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., 2001, *ApJ*, 562, 52
- McDonald P., Seljak U., Cen R., Bode P., Ostriker J. P., 2005, *MNRAS*, 360, 1471
- McDonald P. et al., 2006, *ApJS*, 163, 80
- McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C.-A., 2009, *ApJ*, 694, 842
- McQuinn M., Hernquist L., Lidz A., Zaldarriaga M., 2011, *MNRAS*, 415, 977
- Meiksin A., Bryan G., Machacek M., 2001, *MNRAS*, 327, 296
- Miralda-Escudé J., Cen R., Ostriker J. P., Rauch M., 1996, *ApJ*, 471, 582
- Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1
- Mortlock D. J. et al., 2011, *Nat*, 474, 616
- Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, *MNRAS*, 394, 1812
- Petitjean P., Webb J. K., Rauch M., Carswell R. F., Lanzetta K., 1993, *MNRAS*, 262, 499
- Petitjean P., Mucket J. P., Kates R. E., 1995, *A&A*, 295, L9
- Puchwein E., Pfrommer C., Springel V., Broderick A. E., Chang P., 2012, *MNRAS*, 423, 149P
- Rauch M., 1998, *ARA&A*, 36, 267
- Rauch M. et al., 1997, *ApJ*, 489, 7
- Ricotti M., Gnedin N. Y., Shull J. M., 2000, *ApJ*, 534, 41
- Rollinde E., Petitjean P., Pichon C., Colombi S., Aracil B., D'Odorico V., Haehnelt M. G., 2003, *MNRAS*, 341, 1279
- Rollinde E., Srianand R., Theuns T., Petitjean P., Chand H., 2005, *MNRAS*, 361, 1015
- Schaye J., 2001, *ApJ*, 559, 507
- Schaye J., Dalla Vecchia C., 2008, *MNRAS*, 383, 1210
- Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, *MNRAS*, 318, 817
- Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, *ApJ*, 596, 768
- Schaye J. et al., 2010, *MNRAS*, 402, 1536
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V. et al., 2005, *Nat*, 435, 629
- Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, *MNRAS*, 301, 478
- Theuns T., Schaye J., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell B., 2002a, *ApJ*, 567, L103
- Theuns T., Viel M., Kay S., Schaye J., Carswell R. F., Tzanavaris P., 2002b, *ApJ*, 578, L5
- Theuns T., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell R. F., 2002c, *MNRAS*, 332, 367
- Tytler D., Paschos P., Kirkman D., Norman M. L., Jena T., 2009, *MNRAS*, 393, 723
- Viel M., Haehnelt M. G., 2006, *MNRAS*, 365, 231
- Viel M., Haehnelt M. G., Carswell R. F., Kim T.-S., 2004, *MNRAS*, 349, L33
- Viel M., Bolton J. S., Haehnelt M. G., 2009, *MNRAS*, 399, L39
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tomatore L., 2009b, *MNRAS*, 399, 574
- Zhang Y., Anninos P., Norman M. L., 1995, *ApJ*, 453, L57

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.